Int J Legal Med (2010) 124:617–622
DOI 10.1007/s00414-010-0506-9

ORIGINAL ARTICLE

Developing equine mtDNA profiling for forensic application

Susan M.R. Gurney · Sandra Schneider · René Pflugradt · Elizabeth Barrett · Anna Catharina Forster · Bernd Brinkmann · Thomas Jansen · Peter Forster

Received: 3 July 2010 / Accepted: 16 August 2010 / Published online: 10 September 2010
© Springer-Verlag 2010

Abstract Horse mtDNA profiling can be useful in forensic work investigating degraded samples, hair shafts or highly dilute samples. Degraded DNA often does not allow sequencing of fragments longer than 200 nucleotides. In this study we therefore search for the most discriminatory sections within the hypervariable horse mtDNA control region. Among a random sample of 39 horses, 32 different sequences were identified in a stretch of 921 nucleotides. The sequences were assigned to the published mtDNA types A–G, and to a newly labelled minor type H. The random match probability within the analysed samples is 3.61%, and the average pairwise sequence difference is 15 nucleotides. In a "sliding window" analysis of 200-nucleotide sections of the mtDNA control region, we find that the known repetitive central motif divides the mtDNA control region into a highly diverse segment and a markedly less discriminatory segment.

Keywords Horse · Mitochondrial · Sequence · Control region · D-loop · Genetic · Evolution · Breed

Susan M.R. Gurney and Sandra Schneider contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-010-0506-9) contains supplementary material, which is available to authorized users.

S. M.R. Gurney · B. Brinkmann · P. Forster (✉)
Institute of Forensic Genetics,
48161 Münster, Germany
e-mail: pf223@hermes.cam.ac.uk

S. M.R. Gurney
Institute of Continuing Education, University of Cambridge,
Madingley Hall,
Cambridge CB23 8AQ, UK

S. Schneider
Department of Genetics, University of Cambridge,
Cambridge CB2 3EH, UK

R. Pflugradt · T. Jansen
Certagen GmbH,
Marie-Curie-Str 1,
53359 Rheinbach, Germany

R. Pflugradt
Institute of Legal Medicine, University Hospital Freiburg,
Albertstraße 9,
79104 Freiburg, Germany

E. Barrett
Kenton,
Suffolk IP14 6JS, UK

A. C. Forster
Russee,
24111 Kiel, Germany

P. Forster
Murray Edwards College, University of Cambridge,
Cambridge CB3 0DF, UK

P. Forster
Cambridge Society for the Application of Research, Churchill College,
Cambridge CB3 0DS, UK

P. Forster
Genetic Ancestor Ltd.,
PO Box 503, Cambridge CB1 0AN, UK

Introduction

The mtDNA of horses is maternally inherited, haploid and its genes do not recombine [1]. Excluding new mutations, mtDNA sequences are therefore identical for all matrilineally related horses. The equine mtDNA has an approximate length of 16,660 bp and consists of a coding region, which contains 37 genes [2, 3], and a non-coding region, called the control region, which is approximately 1,192-bp long and contains the displacement loop, see Fig. 1. The control region contains a repetitive region consisting of 2–29 octamer repeats which is prone to length heteroplasmy [4].

Forensic animal DNA profiling is a developing subject in the literature [5–10]. For horses, potential areas for casework includes (a) situations where a horse can be a perpetrator and its identification becomes necessary when it has been involved in an attack on a person, when causing an accident or damaging property, (b) doping in the racing industry, where a horse owner found guilty of doping a horse might invoke laboratory sample confusion in the urine samples analysed by the toxicologist and (c) mtDNA profiling in horse theft cases, especially where only degraded remains, hairs or a distantly and matrilineally related horse are available as sample material. With regard to horse theft, about one in every 6,000–100,000 horses in Western countries are reported, on leading websites, as being stolen at any given time (Table 1). In Germany, pedigree stallions for breeding cost on average $80,000, with the maximum amount paid for in 2008 being $1.5 million [11].

Horse mtDNA profiling would be particularly useful for degraded samples, hair shafts or highly dilute samples, due to the much higher copy number of mtDNA over nuclear DNA. Ideally, the whole mtDNA control region could be amplified to ensure a high genetic resolution. In practice however, the repetitive region can cause amplification problems, and degraded DNA samples often allow the amplification of only short fragments up to 200-bp long [12, 13]. This study therefore sets out to optimise the profiling of horse mtDNA for forensic purposes.

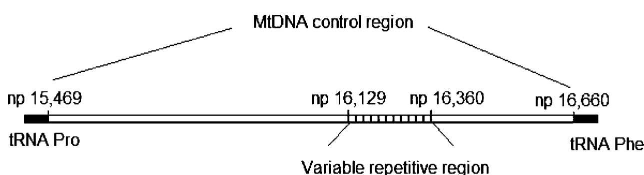

Fig. 1 Horse mtDNA control region. The control region of the horse mtDNA is a non-coding DNA stretch flanked by the tRNA genes for proline and phenylalanine. It is 1,192-bp long (from nps 15,469–16,660) according to the reference sequence [4]. However the length can vary individually due to a repetitive region which consists of around 2–29 repeats of the octamer motif 5′-GTGCACCT-3′, with the published reference sequence representing 29 repeats. In this study, we have also sequenced part of the tRNA-Phe gene (nps 1–70) and into (nps 71–121) the 12S rRNA gene

Material and methods

DNA samples

Tail or mane hair-roots were sampled from 38 unrelated horses representing 11 breeds: one Algerian Barb (sampled in Germany, ancestor from Algeria), one Tunisian Barb (sampled in Germany, ancestor from Tunisia), one Moroccan Barb (sampled in Germany, ancestor from Morocco) five Caspians from the south-eastern shore of the Caspian Sea, four Caspians from the south-western shore of the Caspian Sea, two Dulmeners (sampled in Germany), five Hutzuls (sampled in Romania), five Kurdish horses (sampled in Kurdistan), one Kalmuck (sampled in Russia), six Mongolians (sampled in Mongolia), three Turkomans (sampled in the border area of Turkmenistan/Iran/Afghanistan), one Welsh Cob (sampled in Suffolk, ancestor from Wales), one Welsh Pony (purchased in the Midlands, England), and two Yabous (sampled in the border area of Turkmenistan/Iran/Afghanistan). The published Swedish horse reference sequence [4] was added for the statistical analyses.

DNA extraction, amplification and sequencing

For each horse, total DNA was extracted from five hair roots as published by Allen et al. [14]: lengths of 1 to 2 cm of five hairs with roots incubated for 3 h at 56°C in 200 µL extraction buffer (50 mM KCl, 10 mM Tris-HCl at pH 8.3, 1.5 mM $MgCl_2$, 0.45% Triton X100, 0.45% Tween 20)

Table 1 Horse theft in Europe and America

Country	Horse population[a]	Reported stolen[b]
UK	0.4 million	172[c]
Germany	0.5 million	85[d]
USA	9.5 million	95[e]

[a] Food and Agriculture Organization of the United Nations (FAOSTAT), values for 2008. For the UK, the National Equine Database currently (Aug 2010) has a higher number as it holds 951,320 records of horses with known owners. N.B. not all owners may have registered, and some owners may have omitted to report deceased horses

[b] Reported stolen in the period January 2006 to March 2010 and not recovered by 10 March 2010

[c] Stolen Horse Register website: http://www.stolenhorseregister.com

[d] R. Libowski, pers. comm., and Horse Theft website: http://www.pferdediebstahl.de

[e] Stolen Horse International website: http://www.netposse.com/stolenmissing/stolenhorses.htm

with 2.5 μL proteinase K (20 mg/mL) and 35 mM DTT. Proteinase K was subsequently inactivated by incubating the samples at 95°C for 10 min.

The entire mtDNA control region in each horse was amplified by PCR as two overlapping amplicons covering nucleotide positions (nps) 15,425–16,086 and nps 15,987–00196, as follows.

For the first section (nps 15,425–16,086), published primers were used: 5′-GTAAAACGA CGGCCAGT-ACCATCAACACCCAAAGC-3′ [15] and 5′-TGGTTGCTGATGCGGA-3′ [16]. The PCR was performed in a Touchgene gradient thermocycler (Techne TC-412, Bibby Scientific Limited, Staffordshire, UK) in a total volume of 25 μL with the following reaction conditions: 1 μL DNA template, 2.5 μL 10× Buffer (100 mM Tris HCl, 500 mM KCl, 15 mM $MgCl_2$, 0.1% Tween, 0.1% Triton×100), 1 μL $MgCl_2$ (25 mM stock solution), 2.5 μL of each primer (1 μM stock), 0.2 μM of each dNTP (from a 10 mM stock solution of each dNTP), and 0.2 μL *Taq* polymerase of a 5 U/μL stock (Fermentas, St Leon-Rot, Germany or Qiagen, Hilden, Germany). The PCR cycling conditions were as follows: denaturation at 95°C for 2 min; 35 cycles of 95°C for 30 s, 53°C for 20 s and 72°C for 1 min; final extension step of 72°C for 10 min. The same primers were used for sequencing the amplicons, using the DyeTerminator version 3 kit as specified by the manufacturer (Applied Biosystems, Foster City, USA). The sequences were determined in the range nps 15,480–15,979.

For the second section (nps 15,987–00196), the following primers were used: 5′-GGGATGCTATGACT CAGCTA-3′and 5′-AGTGTGCTTGATACCTGCTC-3′. The expected length of the PCR products is between 654 and 870 bp, depending on the number of repeats within the repetitive region. The PCR buffer and machine was the same as for the amplification of the first section, whereas the PCR conditions differed: 5-min denaturation at 94°C; 35 cycles of 1-min denaturation at 94°C, 45 s annealing at 64.6°C, 1 min 20 s elongation at 72°C; final 5-min elongation step at 72°C, in a reaction volume of 20 μL. The high annealing temperature of 64.6°C was found to be optimal to denature the hairpin secondary structure (Fig. 2) caused by the repetitive region. The sequences were determined in the range nps 16,361–16,660 and nps 1-121.

DNA alignment and diversity analysis

The raw DNA chromatograms were edited in Chromas 2.33 (Technelysium, Tewantin, Australia), the resulting DNA sequences were formatted in DNA Alignment (Fluxus-Technology, Clare, UK), and pairwise genetic distances were calculated in Network 4.510 (Fluxus-Technology, Clare, UK).

Results

The mtDNA sequencing results of 39 horses including the published Swedish reference sequence of Xu and Arnason [4] are shown in Table 2. The nucleotide substitutions are scored against this reference sequence. The sequences were assigned to mtDNA types A–H as a quality control procedure to identify potential errors and specifically in silico recombination between the first and the second section on either side of the repetitive region. This quality control revealed two potential errors in the preliminary analysis, one of which was due to chromatogram quality, and the other was a potential documentation ambiguity. Both problematic sequences were removed, leaving the 39 sequences presented in Table 2. One Welsh Cob is assigned to the new mtDNA type H, which has already been observed for example in Shire Horse JSH233 in Jansen et al. [15], where it represents a minor mtDNA branch.

Among the random sample of 39 horses, 32 different sequences were identified. The majority of sequence types (26) were singletons, five sequence types occurred twice and one sequence type occurred three times. The random match probability for the mtDNA sequence range nps 15,480–00121 (excluding nps 15,980–16,360) within the analysed samples is 3.61%. The pairwise sequence diversity for this stretch of 921 nucleotides is 14.86 nucleotides, i.e., on average, the number of nucleotide differences between a pair of horse mtDNA sequences is nearly 15 nucleotides, irrespective of whether these differences are transitions, transversions, deletions or insertions.

Fig. 2 Secondary structure of the repetitive region in the control region of the horse mtDNA. Because the same motif (5′-GTGCACCT-3′) is present multiple times, the sequence can fold back and form a stable paired double helix [4]. The structure shown here is based on 11 repeats. Experimentally, we found that PCR amplification at standard temperatures (55°C) yielded weak products. Raising the PCR annealing temperature to 65°C increased PCR yield, evidently by denaturing this secondary structure

Table 2 mtDNA sequence variation in 39 randomly selected horses

	15494	15495	15496	15521	15526	15534	15538	15540	15542	15547	15585	15596	15597	15598	15600	15601	15602	15603	15604	15615	15616	15617	15635	15649	15650	15659	15660	15666	15667	15703	15709	15718	15720	15726	15737	15740	15770	15771	15775	15776	15777	15806	15809	15810	15811	15826	15827	15838	15869	15870	15871	15872	15879	15956	15974	16371	16385	16394	16407	16439	16461	16476	16540	16543	16546	16556	16559	16563	16563	16563	16629	77	Type
Ref	T	T	A	G	T	C	A	A	C	C	G	A	A	T	G	T	C	T	G	A	A	T	C	A	A	T	A	G	A	T	C	C	G	G	T	A	C	C	C	T	A	C	A	A	C	A	A	C	C	C	C	A	C	A	C	T	A	A	C	C	A	C	C	T	T	T	C	-	-	-	A	C	A5
B925	.	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	G	.	.	A	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	C	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A1
B926	.	C	.	.	.	.	.	.	T	.	.	.	G	.	.	.	T	.	.	.	.	.	T	.	G	.	.	A	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A1
C6	.	C	.	.	.	.	.	.	T	.	A	.	G	.	.	.	T	.	.	.	.	.	T	.	G	.	.	A	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A1
H7	.	C	.	.	.	.	.	.	T	.	A	.	G	.	.	.	T	.	.	.	.	.	T	.	G	.	.	A	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A1
H8	.	C	.	.	.	.	.	.	T	.	A	.	G	.	.	.	T	.	.	.	.	.	T	.	G	.	.	A	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A1
BC7	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A3
BC8	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A3
T14	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A3
WP13	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A3
BC9	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A4
C3	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A4
C4	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A4
M1000	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	T	G	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A6
M1003	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	T	G	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A6
M1009	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	T	G	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A6
H2	.	C	.	.	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	A	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	.	T	C	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	B2
K13	.	C	.	.	.	.	G	.	.	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	B2
T6	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	T	.	.	.	.	G	.	T	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	C	.	T	.	.	.	G	.	C1
Y3	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	T	.	.	.	.	G	.	T	.	.	.	T	G	.	C	.	.	.	.	.	.	.	A	C	.	T	.	.	.	G	.	C1
BC6	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	T	.	.	.	.	G	T	T	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	C	.	T	.	.	.	G	.	C2
D12	.	C	.	.	.	.	.	.	.	.	A	.	.	.	A	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	T	.	.	.	.	G	T	T	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	C	.	T	.	.	.	G	.	C2
H3	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	T	.	.	.	.	G	T	T	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	C	.	T	.	.	.	G	.	C2
H4	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	T	.	.	.	.	G	T	T	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	C	.	T	.	.	.	G	.	C2
B922	C	C	G	.	.	T	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	G	T	C	.	.	-	.	.	.	.	.	.	.	.	C	C	C	G	.	D1
K1	C	C	G	.	.	T	.	.	.	.	A	.	.	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	G	T	C	.	.	-	.	.	.	.	.	.	.	.	C	.	.	G	.	D2
Km959	C	C	G	.	.	T	.	.	.	.	A	.	.	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	G	T	C	G	.	-	.	.	.	.	.	.	.	.	C	.	.	G	.	D2
BC2	C	C	G	.	.	T	.	.	.	.	A	.	.	.	.	.	T	C	A	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	G	T	C	.	.	-	.	.	.	.	.	.	C	.	.	.	.	G	T	D3
D13	C	C	G	.	.	T	.	.	.	.	A	.	.	.	.	.	T	C	A	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	G	T	C	.	.	-	.	.	.	.	.	.	C	.	C	.	.	G	T	D3
Y1	.	C	.	A	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	E
K14	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	G	C	.	.	A	.	.	.	.	T	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	G	.	C	.	.	-	.	.	.	.	A	.	.	.	C	.	.	G	.	F1
K15	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	G	C	.	.	A	.	.	.	.	T	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	G	.	C	.	.	-	.	.	.	.	A	.	.	.	C	.	.	G	.	F1
K10	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	G	.	.	.	.	.	C	.	.	A	.	.	G	.	T	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	.	.	.	C	.	.	G	.	F2
M1001	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	G	.	T	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	.	.	.	C	.	.	G	.	F2
M989	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	G	.	.	C	.	.	A	A	.	G	.	T	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	G	.	F2
M993	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	G	.	.	C	.	.	A	A	.	G	.	T	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	G	.	F2
C2	.	C	.	.	.	.	.	.	.	T	.	.	.	C	.	.	T	.	.	.	G	.	.	.	.	C	.	.	.	C	.	.	A	.	.	.	T	.	T	C	.	T	.	.	.	.	G	.	.	.	.	.	.	G	T	C	.	.	.	.	.	T	T	A	.	.	.	.	.	.	G	.	G
T7	.	C	.	.	.	.	.	.	.	.	.	.	.	C	.	.	T	.	.	G	G	.	.	.	.	C	.	.	.	C	.	.	A	.	.	.	T	.	T	C	.	T	.	.	.	.	G	.	.	.	T	.	.	G	T	C	.	.	.	.	.	T	T	A	.	.	.	.	.	.	G	.	G
WC3	.	C	.	.	C	.	.	G	T	.	A	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	H

[a] *Ref* denotes the Swedish horse reference sequence in Genbank accession X79547

[b] Sequences are determined in the range nps 15,480–15,979, nps 16,361–16,660 and nps 1–121

[c] Labels designate the breeds as follows: *B* Barb, *BC* southeastern Caspian, *C* southwestern Caspian, *D* Dülmener, *H* Hutzul, *K* Kurdish horse, *Km* Kalmuck, *M* Mongolian, *T* Turkoman, *WC* Welsh Cob, *WP* Welsh Pony, *Y* Yabou

[d] MtDNA types are as defined in refs [15] and [17], with the new type *H* proposed here

To locate, within the mtDNA control region, the most diverse sections of 200 nucleotides, we scanned the control region sequence using a 200-nucleotide window, which we shifted in 20-nucleotide steps. The pairwise diversity value for each of these locations is plotted in Fig. 3.

The most highly variable 200-nucleotide section turns out to be located at nps 15,580–15,779, scoring an average pairwise diversity of 6.50 nucleotide differences. Across the whole analysed region, the mean pairwise difference ranged from a minimum of 0.61 to the maximum of 6.50, compared to the average of 14.86 for the whole analysed sequence range. The two most highly variable non-overlapping 200 bp-sections were firstly, nps 15,580–15,779 (pairwise diversity of 6.50 nucleotides), and secondly, nps 15,780–15,979 (pairwise diversity of 3.91 nucleotides). The post-repetitive sequence stretch does not yield higher diversities than 2.36. Next, we determined the match probability within the sample of the 39 horses for the same 27 sequence sections. These match probabilities are shown in Fig. 4.

As can be seen from Fig. 4, the match probabilities at the beginning of the control region are fairly constant, but then increase significantly after the repetitive region, reaching the highest value as the sliding window progresses into the coding-region tRNAPhe and 12S rRNA genes.

Discussion and conclusions

Profiling horse samples by typing mtDNA is useful when only degraded DNA is available, or in deficiency cases, where material from the horse of interest is not directly available and a matrilineally related horse is sequenced for its mtDNA instead. It is beneficial if the investigator can design his or her PCR primers to target the most discriminatory mtDNA sequence stretch, and to this end

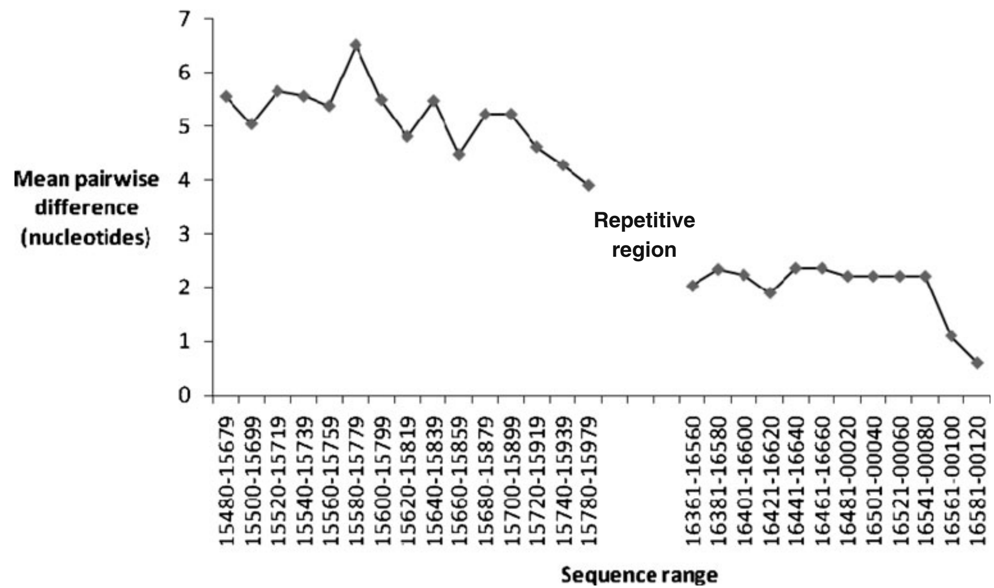

Fig. 3 Determination of highly variable mtDNA sections for forensic application. The mean pairwise difference was analysed for a total of 27 sections (each 200-bp long) and based on mtDNA sequences from 39 horses

we have characterised and analysed 921 nucleotides of the horse mtDNA in a random sample of 39 horses.

We applied two types of diversity analysis to the mtDNA data set we generated, the match probability and the pairwise difference statistic. It is arguably more relevant for forensic casework to use the match probability rather than the pairwise difference statistic; as in general, the question is not how many nucleotides' difference there is between two horse samples, but whether the two samples are identical or not. However, when comparisons between related horses across generations are involved, then there is the theoretical possibility of a mutation, and a single nucleotide difference between two horses might not be considered sufficient for an exclusion. So we performed both calculations in a "sliding window" analysis, and found the results to be compatible, although the pairwise difference diversity predictably allows a more sensitive comparison of the genetic diversity from section to section.

We found that the pairwise difference diversity of mtDNA sections declines by a factor of three when progressing along the mtDNA control region, from the tRNAPro gene to the tRNAPhe gene. Using the pairwise difference statistics, the most highly variable 200-nucleotide section turned out to be located at nps 15,580–15,779, which the investigator should preferentially amplify when dealing with degraded DNA. If sufficient material is available for one further PCR attempt, the pairwise difference statistic indicates that the investigator should

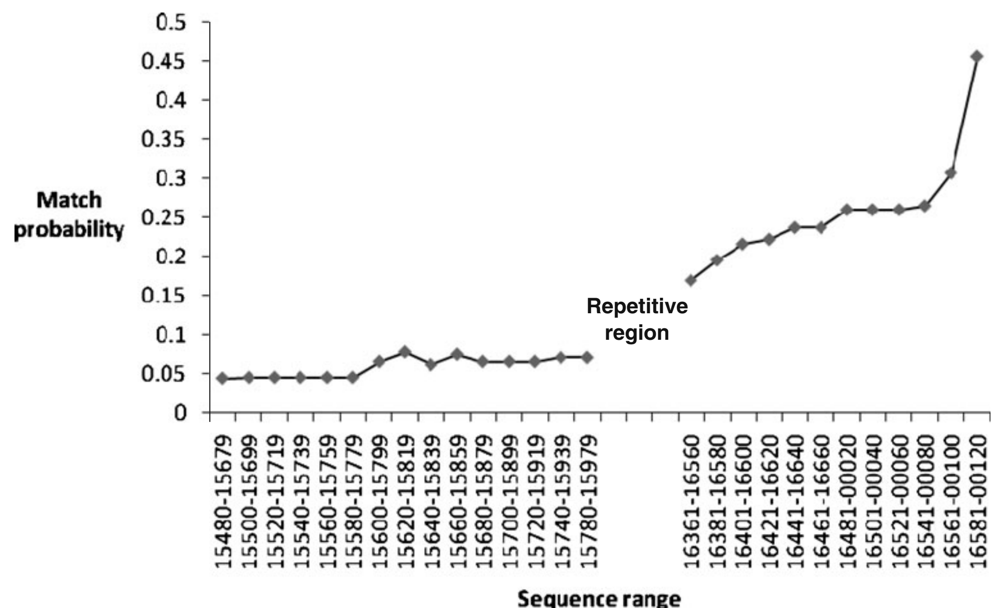

Fig. 4 Determination of highly discriminative mtDNA sections for forensic application. The match probabilities were analysed for a total of 27 sections (each 200-bp long) and based on mtDNA sequences from 39 horses

then preferentially amplify the adjacent 200 bp-section nps 15,780–15,979, and avoid targeting the less discriminatory DNA sections beyond the repetitive region.

An interesting technical observation during this study was that amplification across the repetitive region yielded more products at high temperature of 65°C, and this may be connected to the hairpin secondary DNA structure that is caused by such motif repetitions.

Acknowledgements The authors wish to thank Piotr Miskowicz for providing Eurasian horse samples and Doug Stephens from the National Equine Database Ltd (UK) for information on the UK horse population. This work was performed by SG in partial fulfilment of the requirements for the doctoral degree.

References

1. Bowling AT, Del Valle A, Bowling M (2000) A pedigree-based study of mitochondrial D-loop DNA sequence variation among Arabian horses. Anim Genet 31:1–7
2. Wolstenholme DR (1992) Genetic novelties in mitochondrial genomes of multicellular animals. Curr Opin Genet Dev 2:918–925
3. Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780
4. Xu X, Arnason U (1994) The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. Gene 148:357–362
5. Cooper JE, Cooper ME (2008) Forensic veterinary medicine: a rapidly evolving discipline. Forensic Sci Med Pathol 4:75–82
6. Eichmann C, Parson W (2007) Molecular characterization of the canine mitochondrial DNA control region for forensic applications. Int J Leg Med 121:411–416
7. Imaizumi K, Akutsu T, Miyasaka S, Yoshino M (2007) Development of species identification tests targeting the 16 S ribosomal RNA coding region in mitochondrial DNA. Int J Leg Med 121:184–191
8. Nakamura H, Muro T, Imamura S, Yuasa I (2009) Forensic species identification based on size variation of mitochondrial DNA hypervariable regions. Int J Leg Med 123:177–184
9. Schneider PM, Seo Y, Rittner C (1999) Forensic mtDNA hair analysis excludes a dog from having caused a traffic accident. Int J Leg Med 112:315–316
10. Tsuji A, Ishiko A, Kimura H, Nurimoto M, Kudo K, Ikeda N (2008) Unusual death of a baby: a dog attack and confirmation using human and canine STRs. Int J Leg Med 122:59–62
11. Deutsche Reiterliche Vereinigung (2008) Jahresbericht 2008 der Deutschen Reiterlichen Vereinigung e.V. Bundesverband für Pferdesport und Pferdezucht, Warendorf, p 425
12. Schneider PM, Bender K, Mayr WR, Parson W, Hoste B, Decorte R, Cordonnier J, Vanek D, Morling N, Karjalainen M, Marie-Paule Carlotti C, Sabatier M, Hohoff C, Schmitter H, Pflug W, Wenzel R, Patzelt D, Lessig R, Dobrowolski P, O'Donnell G, Garafano L, Dobosz M, De Knijff P, Mevag B, Pawlowski R, Gusmão L, Conceicao Vide M, Alonso Alonso A, García Fernández O, Sanz Nicolás P, Kihlgreen A, Bär W, Meier V, Teyssier A, Coquoz R, Brandt C, Germann U, Gill P, Hallett J, Greenhalgh M (2004) STR analysis of artificially degraded DNA—results of a collaborative European exercise. Forensic Sci Int 139:123–134
13. Green RE, Briggs AW, Krause J, Prüfer K, Burbano HA, Siebauer M, Lachmann M, Pääbo S (2009) The Neandertal genome and ancient DNA authenticity. EMBO J 28:2494–2502
14. Allen M, Engström AS, Meyers S, Handt O, Saldeen T, von Haeseler A, Pääbo S, Gyllensten U (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J Forensic Sci 43:453–464
15. Jansen T, Forster P, Levine MA, Oelke H, Hurles M, Renfrew C, Weber J, Olek K (2002) Mitochondrial DNA and the origins of the domestic horse. Proc Natl Acad Sci USA 99:10905–10910
16. McGahern AM, Edwards CJ, Bower MA, Heffernan A, Park SD, Brophy PO, Bradley DG, MacHugh DE, Hill EW (2006) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. Anim Genet 37:498–502
17. Vilà C, Leonard JA, Götherstöm A, Marklund S, Sandberg K, Lindén K, Wayne RK, Ellegren H (2001) Widespread origins of domestic horse lineages. Science 291:474–477